Lessons for Today and Tomorrow from Yesterday - the Structure of Alginic Acid

Struther Arnott#,*, Wen Bian*, R. Chandrasekaran*, Bridget Ryan Manis*

The Institute for Cancer Research, Chester Beatty Laboratories, 237 Fulham Road, London SW3 6JB, England.

* Whistler Center for Carbohydrate Research, Purdue University, West Lafayette, Indiana 47907, USA.

We describe a re-determination by X-ray diffraction analysis of the structures of polymannuronic and polyguluronic acids in oriented, polycrystalline fibers and thereby provide a more accurate visualization of the role of divalent cations in alginate gels.

The Context of this Study.

Fibrous substances are important in many biological contexts and their successful exploitation by *Homo sapiens* for food and textured materials had a millenia-long history before it was even thinkable to rationalize their valuable properties in molecular terms. Profitable technologies can evolve without accurate information about molecular or higher structure and even despite inaccurate information: there is no accurate fibre structure in the literature for, say, crystalline rubber! This is not to say that technologies cannot profit from detailed molecular structures in those cases where there is indeed a close and manipulatable relationship between their functional properties and their atomic-level structures. Our predicament is that revealing structures at this level gets to be tantalizingly difficult in many fibrous biological systems where diffraction data are sparse below 2.5 Å. Credible atomic detail therefore comes at a cost more than most fibre diffractionists have been prepared to invest. Poor-mouthing their own good data becomes a routine justification for failure to make this investment and to go on to fully complete and authenticate provisional structures. The published record thereby accumulates too many incomplete and flawed structures and consolidates a needlessly baleful impression of the power of fibre diffraction analysis and the worth of its products. We will use the case of a gel-forming polysaccharide system important in the food and pharmaceutical industries to illustrate the potentially high information content of conventional, film-recorded, fibre diffraction patterns subjected to linked-atom least-squares analysis [1,2] guided by ancillary Fourier syntheses of electron density. In such analyses the polymer framework can, with care, be defined with confidence and precision. Ancillary cations and water molecules justified by the diffraction data then can be added and function rationalised in atomic terms.

The Nature of Alginic Acid.

Alginic acid essentially has a very simple primary structure but its useful properties depend not only on its secondary structure(s) but also on the manner with which small components such as water and ions interact with these polymeric structures. Its biological synthesis occurs via poly-1,4-β-D-mannuropyranosylic acid (M). This polymonosaccharide, mannuronan, then undergoes post-polymeric modification by a C5-epimerase which introduces blocks of poly-1,4-α-L-guluropyranosylic acid (G or guluronan) and regions of mixed M and G composition between the M and G blocks. Alginates with higher content of G show a greater affinity for Ca^{++} ions and form stronger, harder gels. Spinning oriented fibres from such gels has not been very successful but they have been shown to contain the same molecular conformations as the acid forms which can be obtained in uniaxially oriented, polycrystalline form [3,4]. *Faute de mieux* we have had to rely on the data obtained from these pioneering experiments whose good-quality fibre patterns have not been bettered. It is a more common experience in fibre diffraction analyses than in macromolecular crystallography that one has to rely on a unique example of successful art in producing felicitously well-ordered specimens and in this case using the original data has the additional merit of highlighting what could and should have been done at the time of their collection when the same analytical tools that we have used were all available [1].

The X-ray Data and Structural Variables.

We have reviewed the structures of both mannuronan and guluronan. Each in the hands of the original skilled investigators provided what we would regard as rather good fibre diffraction patterns. The numbers of reflections are not large - 44 in the case

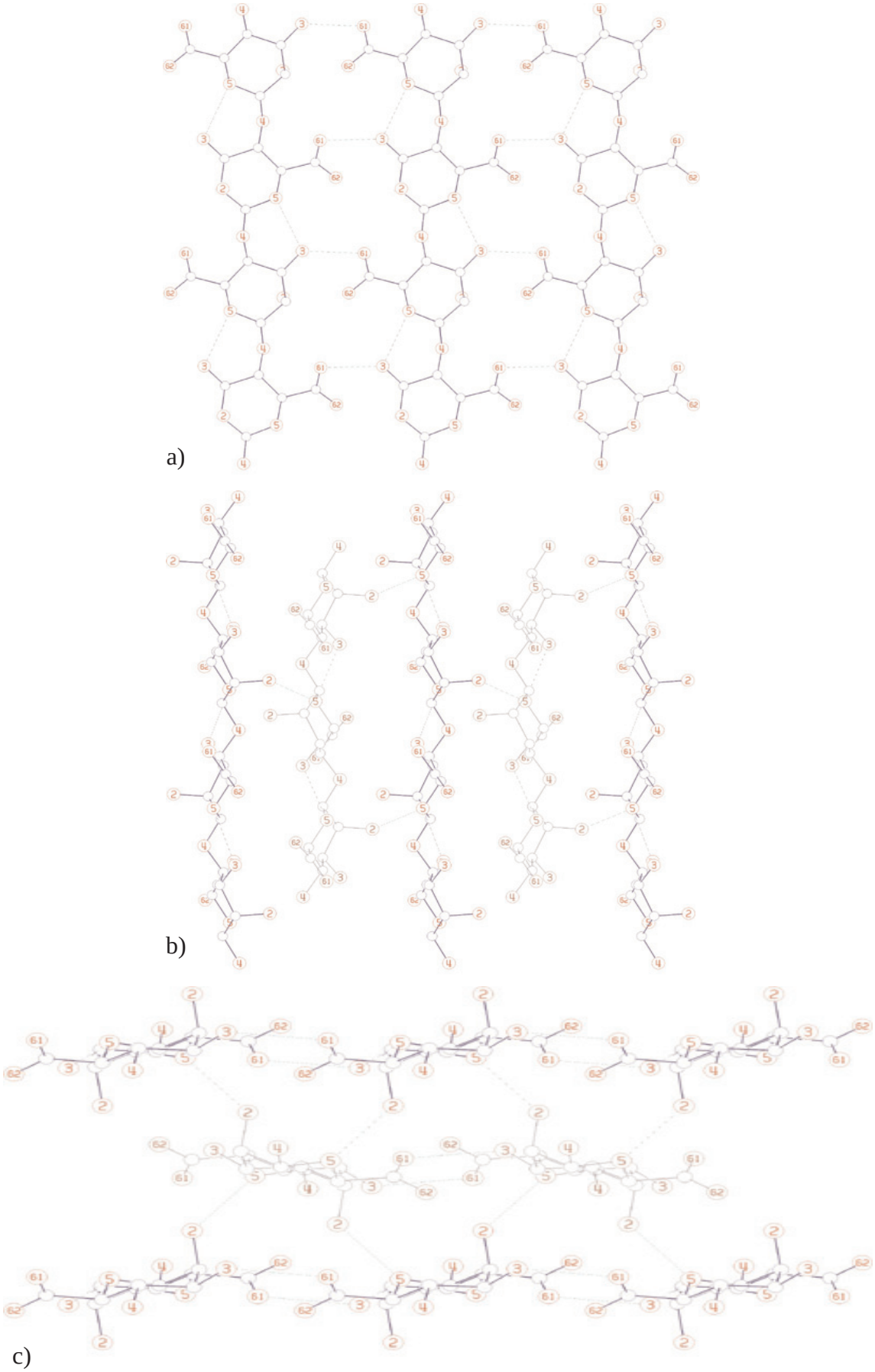

Figure 1: Various views of the crystal structure of mannuronan: (a) the **bc** face of four unit cells, **c** is vertical and the length of the disaccharide is **c** = 10.4 Å, the polysaccharide chains are flat ribbons with O3···O5 H-bonds stabilising the glycosidic linkages, **b** is horizontal and the molecular ribbons are spaced at intervals **b** = 7.6 Å and linked by O3···O61 hydrogen bonds; (b) the **b** axis projection, **c** is again vertical and the horizontal **a** spacing (between parallel molecules) is **a** = 8.6 Å which are linked indirectly by the hydrogen bonds between anti-parallel chains; (c) the **c** axis projection with **a** vertical again showing the inter-sheet O2···O5 H-bonding of the antiparallel sheets of parallel molecules with intra-sheet O3···O61 H-bonding.

of mannuronan and 48 in the case of guluronan - but the molecular structures are correspondingly simple: both are 2-fold helices of monosaccharide residues. With their c spacings determining the pitch, the backbone conformations are fixed except for any restrained variation that was allowed to the initially standard bond angles and ring conformations. The only challenging molecular variable to be determined is the carboxyl conformation and even this can be reduced to a few narrow ranges if there are internal H-bondings such as are illustrated in Figs. 1a and 2a. In both structures two antiparallel 2-fold helices pass through each unit cell in such a fashion as to result in $P2_12_12_1$ orthorhombic symmetry. There are therefore only two variable polymer packing parameters, a molecular translation along **c** and an orientation about **c**. This means that in a linked-atom system with fixed standard bond lengths and angles there are only three degrees of structural freedom to be determined with 45 or so intensity data available in each case. Even when final refinements might involve ring bond and conformation angles subject to elastic restraints on their movements and to ring-closure constraints, the data parameter ratio would remain high, (45+12)/(3+12-6) ~ 6. Even the addition of two water molecules, each with three coordinates, but each also with four hydrogen-bonding restraints, would not seriously reduce the satisfactory ratio, (45+12+8)/(3+12-6+6) ~ 4. Nor should it be forgotten that in these compact unit cells the close contacts between neighbouring non-bonded atoms provide around 70 more quasi-data in the least-squares minimization. Altogether it should be impossible with reasonable care not to arrive at an accurate and precise atomic resolution structure.

The Structures of Mannuronan and Guluronan.

We should define at this point what these two structures are now seen to be. Many important details have emerged from our new analysis - whose details in terms of atomic coordinates, lists of structure factors, etc. are available separately [5] - but the gross framework emerged from the earlier investigation by others and our purpose is not to minimize their contribution but to discuss the difficulties encountered in reaching a credible final conclusion.

The biosynthetic relationship of mannuronan and guluronan and their formally similar crystal structures belie the very different secondary and tertiary structures they assume. The mannuronan chains are extended maximally in the **c** direction just like cellulose with a 10.4 Å repeat (Fig. 1a). Intramolecular O3···O5 H-bonds buttress the resulting molecular ribbons which form sheets when they are laid parallel to one another along the **b** direction with a 7.6 Å spacing and stitched together by intermolecular O61···O3 H-bonds. When these sheets are stacked anti-parallel to one another along **a** (Fig. 1b), the repeat distance is 8.6 Å and the interleaved sheets in the stacks are glued together by O2·O5 H-bonds.

The carboxyl O61 hydrogen atoms have a crucial role in this compact acid structure (Fig. 1c) that has no space for guest molecules even as small as water and therefore does not survive salt formation.

The C5 epimerization which changes mannuronan to guluronan results also in the dramatic change in the pyranose ring conformations which are evident in all views of the G structure (Figs. 2a,b,c). The molecular shape is no longer a ribbon and the **c**-axis repeat is reduced from 10.4 to 8.7 Å and a different H-bond, O2···O61, buttresses the glycosidic linkage (Fig. 2a). There is no longer direct H-bonding between polysaccharide chains. Instead a chain of water molecules provides O2···W···O61 bridges between guluronan molecules along the **a** direction, using every other water molecule in the chain. The alternating set of water molecules in the chain provides O3···W···O5 bridges between the polysaccharide chains that are lined up parallel to one another along the **b** direction (Fig. 2b). The images given by this somewhat artificial deconstruction of the overall three-dimensional structure needs to be integrated by the **c**-axis projection of the lattice (Fig. 2c) which shows a more comprehensive picture of parallel chains at the corners of two contiguous unit cells and the central anti-parallel chains completing the orthorhombic array. The water molecules fill up the channels that are diamond-shaped in cross section and run along the **c** direction. Although these water molecules are connected vertically to one another, laterally each also provides a bridge between two parallel chains across the diagonals of the diamond, but none by itself is connected to more than two chains. The picture changes dramatically when water is replaced by Ca^{++} in one water site (Fig. 3). Then the guest molecule becomes the connector of all four guluronan chains bordering the channel: the O5···X···O3 connection acquires an intersecting

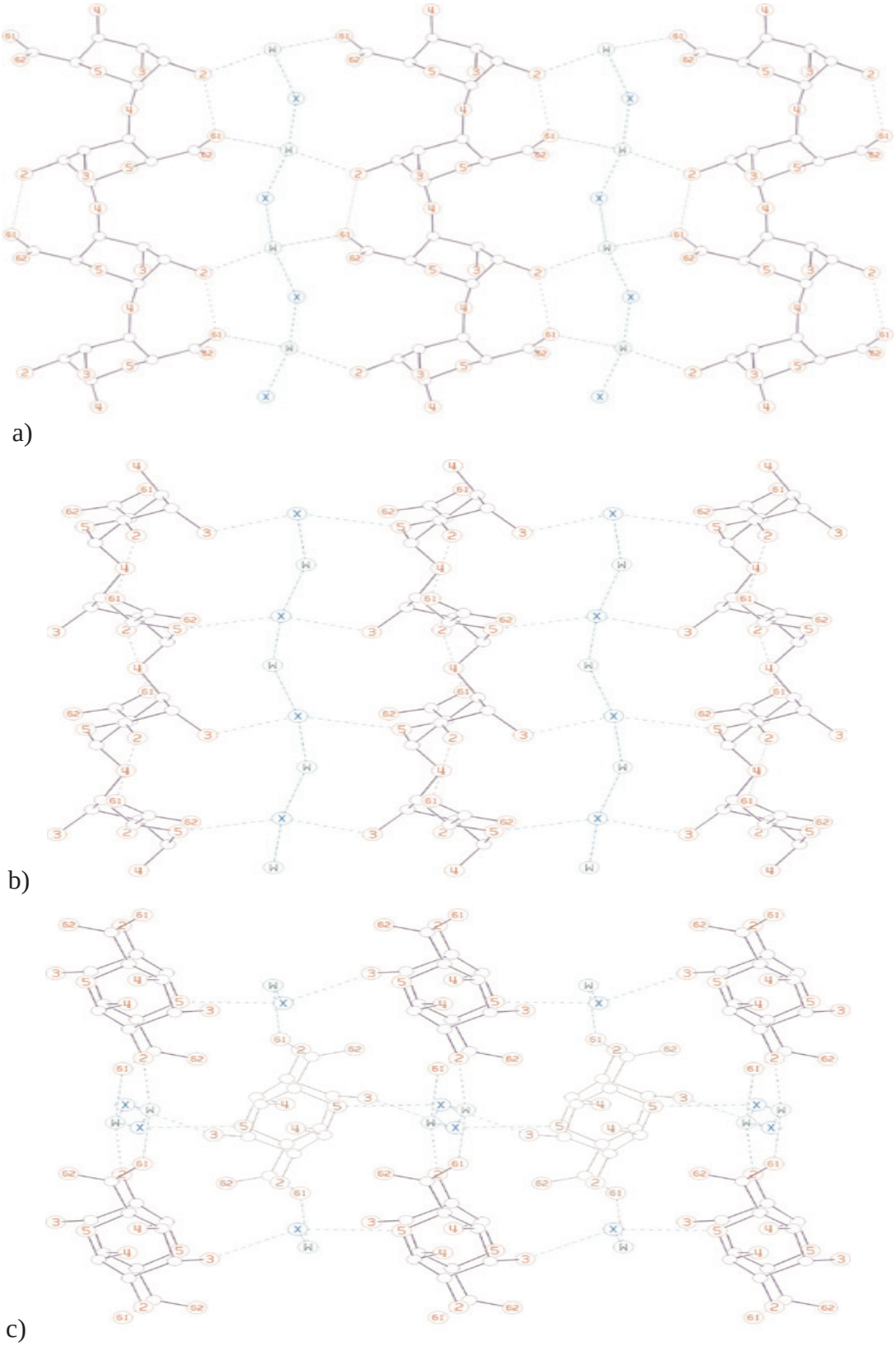

Figure 2: Various views of the crystal structure of hydrated guluronan: (a) the **ac** face of four unit cells, **c** is vertical and the length of the disaccharide repeat is much reduced relative to mannuronan (**c** = 8.7 Å), the intra-chain H-bonds are now O2···O61, the **a** spacing is long (**a** = 10.7 Å) and accommodates a chain of water molecules between the parallel guluronan chains, linking them by O2···W···O61 H-bonds through the set of 21 symmetry related water molecules W; (b) the **bc** face with **c** vertical showing that these parallel guluronan chains are linked by O3···X···O5 H-bonds using the second set of symmetry-related water molecules X, the chain-to-chain spacing **b** = 8.6 Å; (c) the **c** axis projection of two complete unit cells with a vertical, the central diamond shows all the O···water···O bridges that glue the matrix together - the (green) O61···W···O2 bridges vertically and the (blue) O3···X···O5 bridges horizontally.

connection O2···X ···O61 more or less perpendicular to it and, like it, perpendicular also to the vertical string of water molecules. It is not surprising that the resulting gel is strong and hard. Nor is it surprising that Sr^{++} ions promote even tougher gels since the $O\cdots X^{++}$ distances in the polysaccharide matrix would be an even better match with a divalent cation of radius a few tenths of an Å greater than Ca^{++}. All in all the picture of the cation-dependent alginate structure that emerges from this analysis is profoundly different from the egg-box cartoon that was developed by Rees *et al*. [6] in the wake of the earlier imperfect structure.

The Sequence of the X-ray Refinements.

It had not been our intention to embark on a major recapitulation of the structure determination of these two polysaccharide systems. We had in mind only to have on file up-to-date versions of the structures with confirmed stereochemistry and added detail such as the hydrogen atoms omitted in the earlier study. We also wanted to recover an unbiased set of X-ray data with reflections too close to separate grouped together as one composite datum rather than carved up in the same ratio as some calculated model set as had been done in the provisional work. We wanted also to make sure that all reflections perceived to be absent within the field of observed reflections were indeed calculated to be below the observational threshold for any modified structures we might construct. We planned a routine final refinement to reconcile our modestly modified version of the original structures with these equally modestly modified data but began with an attempt to reproduce the original final structures and their structure factors. Unfortunately we discovered that we could in neither case reproduce these (and their R-factors) without improperly doubling the values of our calculated *h*0*l* and 0*kI* reflections. We had thereby uncovered, not for the first time, a not uncommon error in structure analyses involving uniaxially oriented polycrystalline fibres [7]. It can arise from the facile cliche that all diffraction data from fibres are 'cylindrically averaged' versions of the Fourier transform of a molecule or a molecular assembly. This is just not the case for oriented and polycrystalline specimens where one is dealing with the analogue of a rotation diagram of a single crystal where different zones of reflections may have different multiplicities. In the orthorhombic case there is systematic overlap of 0,*k*,*l* and 0,-*k*,*l*, -- only two reflections assumed to have equal amplitudes in this instance -- but for *h*,*k*,*l* there are four reflexions that overlap systematically and also can be assumed to have equal amplitudes: -*h*,*k*,*l* and *h*,-*k*,*l* and -*h*,-*k*,*l* and *h*,*k*,*l*. The important thing to do is to compare observed and calculated quantities that correspond with one another!

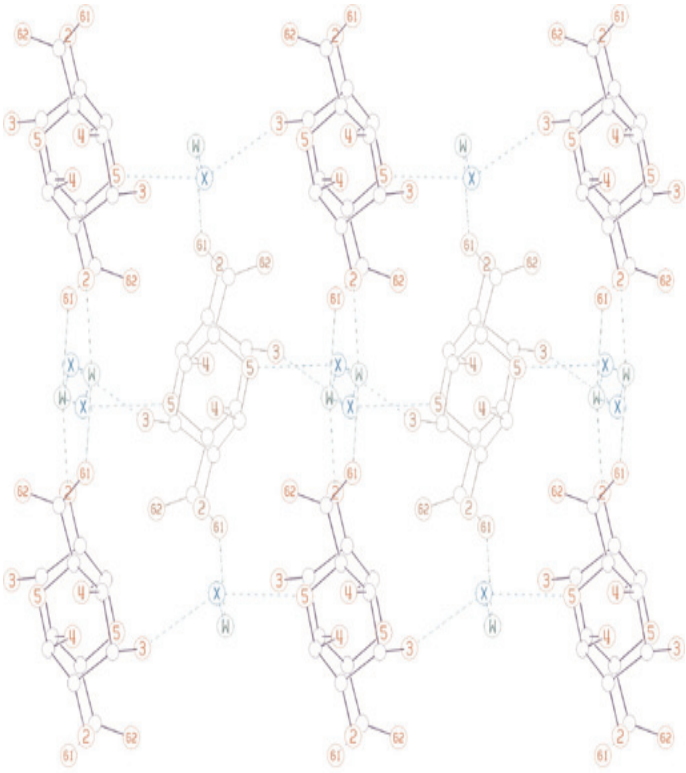

Figure 3: When X is not water but Ca^{++} the resulting octahedral coordination allows this component to make bridges vertically between O2 and O61 as well as horizontally between O3 and O5. Each water coordinated to Ca^{++} continues to participate in the O2···W···O61 bridges shown in Figs. 2a,c but not shown in this illustration.

Correcting the M and G data sets for these fundamental errors immediately raised their already non-descript 'final' *R* values, respectively 0.23 and 0.25, to provocatively higher values, 0.31 and 0.40. We therefore had to go on with new refinements using the corrected data set but starting with the final parameters provided by the previous analyses. In the discussion which follows, we will continue to quote the conventional *R* factor as a more familiar surrogate for the quantities we in fact used to monitor the course of our refinements. The quantity minimized by the LALS process is:

$$\Omega = \Sigma\, \omega\, \Delta F^2 + \Sigma\, \varepsilon\, \Delta\Theta^2 + \Sigma\, \kappa\, \Delta D^2 + \Sigma\, \lambda\, G$$

and the statistical tests [8] of the significance of improvements achieved during our structure analyses were made using its values and/or more conservatively its first component , $\Sigma\omega\Delta F^2$, which is the usual sum of the weighted, squared differences between model and experimental structure amplitudes. The second sum of weighted, squared difference terms involves structure parameters such as bond-angles or ring conformation-angles that have standard values that are explicitly varied during an analysis while tied elastically to the standard values. The third term involves another set of restrained variables (such as hydrogen bond lengths) that may vary elastically from pre-set standard values and are functions of the parameters explicitly varied during the least-squares refinements. The fourth term involves Lagrange Multipliers (λ) and the constraint functions (G) that have all to be made zero by the end of the refinement. Commonly these (exact and not elastic) relationships are used to ensure ring closure, residue connectivity and the like.

For the M structure, refinement of a scale and an overall isotropic temperature factor as well as the packing variables and the carboxyl conformation brought *R* down from 0.31 to 0.21 but changing the domain of the carboxyl conformation so that its refined value ended up at 10^o rather than 114^o reduced *R* still further to 0.16. At this stage difference electron density maps were used to confirm that there were no guest molecules and the analysis was concluded by a refinement that permitted restrained variation of the previously fixed ring conformations and bond angles. This resulted in reducing *R* falling to 0.12.

With the G structure, the same sequence of refinements starting with *R* at 0.40 brought improvement only to 0.37 before reduction could be continued to 0.27 by moving the carboxyl conformation to an alternative domain in this structure also. At this stage, by contrast to the M structure, there was clear evidence from the electron density syntheses of further scattering material centred half way along **a** and about 3 Å from two carboxyl groups. A single water molecule placed near this position resulted in *R* = 0.25, but a pair of water molecules brought *R* down to 0.20. A concluding refinement with relaxed ring conformations ended with *R* = 0.17.

Technical Lessons.

Data. The obligation of diffractionists to publish their data at the same time as the structure(s) derived from them is now well established among crystallographers. The same obligation must be expected of fiber diffractionists. We were fortunate in this analysis that our predecessors had been so scrupulous as to do just this at a time when it was much less common. We were fortunate also that the error in their data was so simple and straightforward to correct. For other diffraction studies where there has been elaborate manipulation of the fundamental experimental observations it may be desirable that these unprocessed data be freely available also. This has to be a serious option in fibre diffraction analyses where many procedures are pioneered in local laboratories but do not become part of some canon recognised by a large community as is the case with, say, protein crystallography.

R Factors. The original, flawed data sets with partially refined structures subjected to limited but unspecified elasticity gave *R* factors of 0.23 and 0.24 respectively for mannuronan and for guluronan (with its one added water molecule per sugar residue). The corrected data sets and the same gross polysaccharide structures refined with fixed bond lengths and angles and fixed sugar rings and no guest molecules gave *R* factors of 0.21 and 0.37 after refining a scale and a temperature factor, two packing parameters and the carboxyl conformations (in wrong domains). We can estimate that adding one or two water molecules to the guluronan structure would have reduced *R* to 0.34 or 0.28 but we know that in the end both mannuronan and guluronan ended up with these *R* factors nearly halved as a result, first, of changing the domains of the carboxyl conformations (a significant 30% reduction of the previous value) and, finally, restrained relaxation of the ring shapes - to anneal or equilibrate, so to speak, the completed structure - provided the final 15% improvement. This also was a significant improvement as determined - as always in our analyses - by a Hamilton test [8]. The final *R* values (0.12 and 0.17) confirm our original good impression of the efforts of the pioneering investigators. Because these values have been subjected to Hamilton Test monitoring, we believe they are not cosmetic and, indeed, the harmonious structures that have emerged as a result of our scrupulous step-wise processes are consistent with this conclusion. The historic data from good laboratories that provided these good indices are therefore deserving of

considerable respect. Processing such data could be improved in simple non-controversial ways. There is evidence that the main systematic errors are due to background and baselines and that very significant improvements in the fit between observed data and their modelled equivalents can be achieved by intruding one or two global parameters to deal with overall background. Nevertheless those of us who deal mainly with biological or other solvated systems in the end will come up against a phenomenon observed, and commented upon [9], by macromolecular crystallographers: *R* factors, even in very favourable cases, are systematically higher for data of resolution > 5 Å than they are for higher resolution data even when substantial allowance has been made for solvent. In many fibre systems, where the data-to-parameter ratios are quite similar although the data are many fewer, a higher proportion of data come from the difficult range so that miniscule values of *R* will be denied to us no matter how sophisticated our data detectors become. In the context of this discussion it has not escaped our notice that the compact, unsolvated mannuronan structure has a rather lower *R* value than the hydrated guluronan structure even although its data-to-parameter ratio is somewhat greater than that of guluronan!

Least-Squares.

We have continued to exploit a modern version of the Linked-Atom Least-Squares technology and eschewed hybridizing it with algorithms that minimize free energy. Both approaches have their drawbacks. Our main concern is to perform a diffraction analysis augmented, minimally, with other information in order to overcome the somewhat limited resolving power of the diffraction data, and not to compensate either for their often small number (since we always restrict the number of allowed variables and their nature appropriately), nor yet for their quality (since it is evident from analyses such as the present ones that the data often are good enough to allow structures to be solved, with confidence and precision, to an accuracy that is within a few tenths of an Ångstrom in atomic positions). To some extent the additional restraint terms in W relate to the potential energy of the structure being determined but they have the advantage of having the right form to be integrated with a least-squares process not least because it can be arranged that in the end the relative weights of all the terms can be harmonized to ensure that all sub-sets have mean squared discrepancies of unity (i.e. $<\omega\Delta F^2> \sim <\varepsilon\Delta\Theta^2> \sim <\kappa\Delta D^2> \sim 1$). This is the condition for a successful least-squares minimisation from which meaningful measures of overall significance can be extracted as well as standard deviations for the values of variables [10].

Optimization.

Optimization has to be a way of life for fiber diffraction analysts. Solution of the phase problem by them most often is dependent on the formulation of initial models. If alternative models can be proposed, a choice between them should not be made on the basis of samples of each type but on the basis of the optimum representatives. A related but distinct problem is the occurrence of false minima where a best fit is obtained with all restraints and constraints satisfied but a better solution exists that can be accessed only by conducting the refinement from a different starting point. The analyses of mannuronan and guluronan each provided examples of this in the matter of their carboxyl conformations. These have two sterically viable domains about 90º apart (and a further two, 180º from each of these, which can be distinguished from the original two only when carboxyl can be distinguished from carboxylate). In both investigations, the earlier workers had the bad luck to opt for conformations in the wrong domain. This we believe was not a result of their incorrect sets of data since we achieved for mannuronan a very satisfactory *R* =0.21 with the corrected data and a refined version of their inaccurate structure. Rather their mistake was to make a premature choice of the conformational domain of the major variable (the carboxyl), before there had been a concerted refinement of all the variables as would occur in a least-squares process. Then they compounded the initial mistake by not giving a further opportunity to the alternative conformation in a second refinement. Such parsimony is inappropriate in fibre diffraction studies: when alternatives exist, significant discrimination can be made only between optimum representatives of each possibility.

Acknowledgements.

This work was supported by the UK Leverhulme Trust and the Industrial Consortium of the Whistler Center for Carbohydrate Research.

References.

[1] Arnott, S. and Wonacott, A.J., *Polymer* (1966) 7, 157-166.
[2] Smith, P.J.C. and Arnott, S., Acta *Crystallogr*. (1978) A34, 3-11.
[3] Atkins, E.D.T., Nieduszinski, I.A., Mackie, W., Parker, K.D. and Smolko, E.E., *Biopolymers* (1973) 12, 1865-1878.
[4] Atkins, E.D.T., Nieduszinski, I.A., Mackie, W., Parker, K.D. and Smolko, E.E., *Biopolymers* (1973) 12, 1879-1887.
[5] Bian, W., Ph.D. Thesis, Purdue University (2001).
[6] Grant, G.T., Morris, E.R., Rees, D.A., Smith, P.J.C. and Thom, D., *FEBS Letters* (1973) 32, 195-198.
[7] Upstill, C., Atkins, E.D.T. and Attwool, P.T., *Int. J. Biol. Macromol.* (1986) 8, 275-288.
[8] Hamilton, W., *Acta Crystallogr.* (1965) 18, 502-510.
[9] Sheldrick, G.M. and Schneider, T.R., *Methods Enzymol.* (1997) 277, 319-343.
[10] Cruickshank, D.W.J., *International Tables for X-ray Crystallography* (1972) Vol. II, 84-98.

Applications of highly constrained molecular modelling scattering curve fits to biologically important proteins

Stephen J. Perkins

Dept of Biochemistry and Molecular Biology, Royal Free and University Medical School, Rowland Hill Street, London NW3 2PF, UK.

Full molecular structures can be extracted from solution scattering analyses of multidomain or oligomeric proteins if the scattering curve fits are constrained by known small structures for the subunits. All the different possible molecular structures are computed, using as constraints any known covalent connections or symmetry features between the subunits. Each model is assessed for steric overlap, radii of gyration, sedimentation coefficient and R-factor. Filtering leaves a small family of good fit models that corresponds to the molecular structure of interest. These structural analyses often provide new biological insights into function.

Introduction

Solution scattering is a diffraction technique used to study overall structures in solution. A sample is irradiated by a collimated, monochromated beam of X-rays or neutrons. The resulting two-dimensional circularly-symmetric diffraction pattern is recorded on a flat area detector system. Radial averaging leads to a one-dimensional scattering curve. Traditionally these curves leads to structural information at a resolution of about 2-4 nm from calculations of the radius of gyration R_G, the cross-sectional R_G (R_{XS}) and the distance distribution function, and the use of spherical harmonics or genetic algorithms provides an overall view of the macromolecule. This approach provides information on overall macromolecular dimensions and molecular weights from the intensity I(0) at zero scattering angle. In distinction to this traditional approach, the utility of solution scattering has been much improved by means of a novel strategy in which molecular structures are derived directly from the scattering curves. This method starts from known molecular structures for subunits within the macromolecule which are used as tight constraints of the scattering data (reviewed in [1-3]).